Title

Draft genome of Haloarcula rubripromontorii strain SL3, a novel halophilic archaeon isolated from the solar salterns of Cabo Rojo, Puerto Rico.

Authors

Sánchez-Nieves, Rubén
Facciotti, Marc
Saavedra-Collado, Sofía
et al.

Publication Date

2016-03-01

DOI

10.1016/j.gdata.2016.02.005

Peer reviewed

Genomics Data 7 (2016) 287–289

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Draft genome of *Haloarcula rubripromontorii* strain SL3, a novel halophilic archaeon isolated from the solar salterns of Cabo Rojo, Puerto Rico

Rubén Sánchez-Nieves [a], Marc Facciotti [b], Sofía Saavedra-Collado [a], Lizbeth Dávila-Santiago [a], Roy Rodríguez-Carrero [a], Rafael Montalvo-Rodríguez [a,*]

[a] Biology Department, University of Puerto Rico, Mayaguez, Box 9000, 00681-9000, Puerto Rico
[b] Biomedical Engineering and Genome Center, 451 Health Sciences Drive, Davis, CA, 95618, United States

ARTICLE INFO

Article history:
Received 1 February 2016
Accepted 5 February 2016
Available online 6 February 2016

Keywords:
Halophilic archaea
Henome sequence
Puerto Rico

ABSTRACT

The genus *Haloarcula* belongs to the family *Halobacteriaceae* which currently has 10 valid species. Here we report the draft genome sequence of strain SL3, a new species within this genus, isolated from the Solar Salterns of Cabo Rojo, Puerto Rico. Genome assembly performed using NGEN Assembler resulted in 18 contigs (N50 = 601,911 bp), the largest of which contains 1,023,775 bp. The genome consists of 3.97 MB and has a GC content of 61.97%. Like all species of *Haloarcula*, the genome encodes heterogeneous copies of the small subunit ribosomal RNA. In addition, the genome includes 6 rRNAs, 48 tRNAs, and 3797 protein coding sequences. Several carbohydrate-active enzymes genes were found, as well as enzymes involved in the dihydroxyacetone processing pathway which are not found in other *Haloarcula* species. The NCBI accession number for this genome is LIUF00000000 and the strain deposit number is CECT9001.

Specifications:	
Organism	*Haloarcula rubripromontorii*
Strain	SL3
Sequencer or array type	MiSeq Systems (Illumina)
Data format	Analyzed
Experimental factors	Microbial strain
Experimental features	Assembled and annotated whole genome
Consent	N/A
Sample source location	Solar Salterns of Cabo Rojo, Puerto Rico 17°57′12″N, 67°11′45″W

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/bioproject/PRJNA293564
Summary:
Genome information of Haloarcula rubripromontorii.

Organism	*Haloarcula rubripromontorii*, Strain SL3
Source	Cabo Rojo, Puerto Rico
Genome Size, Mb	3.97
GC content, %	61.97
tRNA	48
rRNA	6
Protein coding sequences	3797

Solar salterns are hypersaline environments of great historical and commercial importance. These hypersaline aquatic environments are termed thalassohaline indicating that the high salt content is proportional to the concentration of ions found in seawater [1]. These environments have unique ecologies and have become an important resource for the isolation and study of halophilic archaea [7,8]. These organisms are not only key models for developing our understanding of the archaeal domain but may also contribute to the function of the salterns themselves and have potential application outside of their salty homes. For instance, the presence of microbial pigments, like carotenoids, in this system may help to increase the evaporation rate of seawater [2,3]. In addition, some microbes isolated from solar salterns have been found to possess properties that are important for biotechnological applications [4,5,6].

The microbial life in the solar salterns of Cabo Rojo, Puerto Rico has been studied for over 25 years, and several new species have been isolated and described [7,8]. Recently, strain SL3 was isolated from these salterns (17°57′12″N, 67°11′45″W; Fall '14) by selection on agar plates with glycerol as the sole carbon and energy source. Analysis of the 16S rRNA gene revealed that this strain belonged to the genus *Haloarcula* which currently contains ten formally described species [9]. Phylogenetic analyses using the *rpoB*, *ppsA*, and *atpB* genes and other physiological characteristics demonstrate that strain SL3 is a new species within this genus. The name *Haloarcula rubripromontorii* is proposed and a formal taxonomic description is in progress. Strain SL3 has been deposited in CECT with the accession number 9001.

* Corresponding author.
E-mail address: rafael.montalvo@upr.edu (R. Montalvo-Rodríguez).

http://dx.doi.org/10.1016/j.gdata.2016.02.005

The draft genome consists of 3,970,989 bp in length with G + C content of 61.97%. The genome was predicted to include 3797 open reading frames (ORFs) and 6 rRNA (5S(3), 16S(2), 23S(1)) and 49 tRNA genes. Based on RAST functional categories (Fig. 1) (http://rast.nmpdr.org/), the largest functional category (RAST subsystems) of ORFs belong to carbohydrate metabolism with 311 annotated genes. A total of 115 carbohydrate-active enzymes were annotated from the genome and categorized (Table 1) by dbCAN (http://csbl.bmb.uga.edu/dbCAN/). When compared to the genomes of other members of the genus *Haloarcula*, it was observed that strain SL3 possess genes which translate into subunits of the phophoenolpyruvate-dihydroxyacetone phosphotransferase, these genes are not found in any other species of *Haloarcula*. This enzyme is induced as part of glycolysis when cells are grown using dihydroxyacetone as a sole carbon source [10]. In contrast, proteins associated with glycerol-3-phosphate (G3P) ABC transport were not found in *Haloarcula rubripromontorii*, yet they exist within all other *Haloarcula* species. Glycerol is a precursor to G3P in lipid biosynthesis [11]. The absence of a G3P transport might not be necessary if glycerol is readily available for lipid biosynthesis.

2. Experimental design

Strain SL3 was isolated in media containing sterile pond water from the salterns which was diluted to 20% NaCl (*w*/*v*) where yeast extract (5 g/L), and glycerol (5 ml/L) were added. Agar was used as solidifying agent (20 g/L). Genomic DNA extraction was performed using a Promega Wizard® Genomic DNA Purification Kit. The DNA sample was sequenced at the Molecular Research Lab (MR DNA) facility in Shallowater, TX, USA. The Nextera DNA Sample preparation kit (Illumina) was used following the manufacturer's instructions to prepare the genomic library. The initial DNA concentration was determined using the Qubit® dsDNA HS Assay Kit (Life Technologies). To achieve the recommended DNA input of 50 ng, samples were diluted accordingly at a concentration of 2.5 ng/μL. Then samples underwent fragmentation and the addition of adapter sequences. These adapters are utilized during a limited-cycle (5 cycles) PCR in which unique index was added to the sample. After the library was prepared, the final concentration was measured with the Qubit® dsDNA HS Assay Kit (Life Technologies), and the Agilent 2100 Bioanalyzer (Agilent Technologies) was utilized to determine the average library size. The library was then pooled in equimolar ratios of 2 nM, and 12 pM of the library pool was sequenced paired end for 600 cycles using the MiSeq system (Illumina). Assembly was performed using the NGEN Assembler and resulted in 18 contigs (N50 = 601,911 bp), the largest of which included 1,023,775 bp. Protein coding sequences were predicted using RAST (Rapid Annotation Subsystem Technology) [12], ribosomal RNA genes were detected using RNAmmer 1.2 server [13], and tRNA genes were detected using ARAGORN [14]. Carbohydrate-active enzymes were annotated by dbCAN [15].

Conflict of interest

The authors declare that not conflict of interest exist about the work published in this paper.

Acknowledgement

This study was supported by Howard Hughes Medical Institute (HHMI) "Enhancing Advanced Educational Opportunities in STEM Fields for Minority Students at UPRM" Grant Number 52007566. The

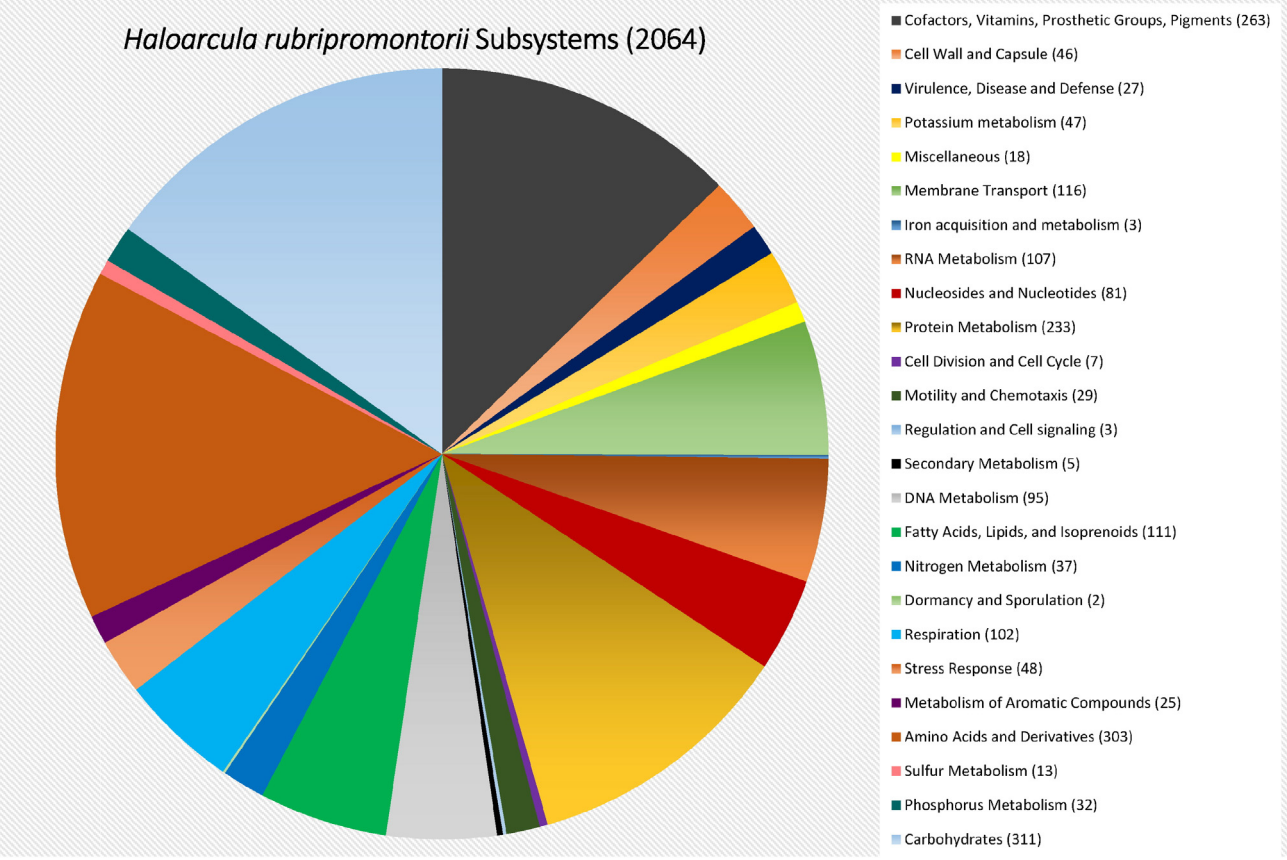

Fig. 1. The subsystem category distribution of strain SL3 (*Haloarcula rubripromontorii*). The chart represents the coverage of proteins which were grouped into subsystems. Each section represents a subsystem and the number of proteins within that subsystem. A total of 2064 proteins were categorized within these subsystems. This chart was generated by RAST (Rapid Annotation System Technology).

Table 1
CAZyme annotation of strain SL3 (*H. rubripromontorii*).

Class	Family	Number of members
Auxiliary Activity (AA)	2	6
	3	2
	6	1
	7	1
Carbohydrate-Binding Module (CBM)	6	3
	13	1
	16	1
	35	2
	40	1
	44	3
Carbohydrate Esterase (CE)	1	4
	4	3
	8	2
	10	3
	14	1
Dockerin		2
Glycoside Hydrolase (GH)	2	1
	3	2
	13	6
	15	3
	32	1
	36	1
	42	1
	68	1
	74	3
	77	1
	97	1
	99	1
	109	8
	120	1
Glycosyl Transferase (GT)	1	1
	2	12
	4	21
	19	1
	20	1
	40	1
	66	3
	75	1
	81	1
	83	3
	94	1
Polysaccharide Lyase (PL)	5	1
	12	1

authors would like to thank members of the Extremophile Laboratory at UPRM for their help.

References

[1] W.D. Williams, 1. Inland salt lakes: An introduction. Hydrobiologia 81 (1) (1981) 1–14.
[2] A. Oren, F. Rodríguez-Valera, The contribution of halophilic bacteria to the red coloration of saltern crystallizer ponds (1). *FEMS Microbiol. Ecol.* 36 (2–3) (2001) 123–130.
[3] A. Oren, Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. Front. Microbiol. 4 (315) (2013) 1–6.
[4] A.G. Jones, C.M. Ewing, M.V. Melvin, Biotechnology of solar saltfields. Salt Lakes 406 (1981) 391–406.
[5] C.D. Litchfield, Potential for industrial products from the halophilic Archaea. J. Ind. Microbiol. Biotechnol. 38 (10) (2011) 1635–1647.
[6] A. Oren, The ecology of *Dunaliella* in high-salt environments. *J. Biol. Res. (Thessaloniki)* 21 (1) (2014) 23.
[7] R. Montalvo-Rodríguez, R.H. Vreeland, A. Oren, M. Kessel, C. Betancourt, J. López-Garriga, *Halogeometricum borinquense* gen. nov., sp. nov., a Novel Halophilic Archaeon from Puerto Rico. Int. J. Syst. Bacteriol. 48 (1998) 1305–1312.
[8] R. Montalvo-Rodríguez, J. López-Garriga, R.H. Vreeland, A. Oren, A. Ventosa, M. Kamekura, *Haloterrigena thermotolerans* sp. nov., a halophilic archaeon from Puerto Rico. *Int. J. Syst. Evol. Microbiol.* 50 (2000) 1065–1071.
[9] A.C. Parte, LPSN–list of prokaryotic names with standing in nomenclature. Nucleic Acids Res. 42 (Database issue) (2014) D613–D616.
[10] R.Z. Jin, E.C. Lin, An inducible phosphoenolpyruvate: dihydroxyacetone phosphotransferase system in *Escherichia coli*. J. Gen. Microbiol. 130 (1) (1984) 83–88.
[11] M. Falb, K. Müller, L. Königsmaier, T. Oberwinkler, P. Horn, S. von Gronau, D. Oesterhelt, Metabolism of halophilic archaea. Extremophiles 12 (2) (2008) 177–196.
[12] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, R. Stevens, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42 (Database issue) (2014) D206–D214.
[13] K. Lagesen, P.F. Hallin, E. Rødland, H.H. Stærfeldt, T. Rognes, D.W. Ussery, RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. 35 (9) (2007) 3100–3108.
[14] D. Laslett, B. Canback, ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32 (1) (2004) 11–16.
[15] Y. Yin, X. Mao, J.C. Yang, X. Chen, F. Mao, Y. Xu, dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40 (Web Server issue) (2012) W445–W451.